

Color appearance: Properties of the uniform appearance diagram derived from hue and saturation scaling

ISRAEL ABRAMOV

Brooklyn College, CUNY, Brooklyn, New York
and Graduate Center, CUNY, New York, New York

JAMES GORDON

Hunter College, CUNY, New York, New York
and Graduate Center, CUNY, New York, New York

AND

HOOVER CHAN

Schools of the Sacred Heart, San Francisco, California
and Eastwind Associates, San Francisco, California

Color appearance can be specified by a procedure of direct hue and saturation scaling. The scaling data can be represented on a 2-D color space termed a *uniform appearance diagram* (UAD). The orthogonal and bipolar axes of the UAD are based on the four unique hue sensations, which are blue–yellow and green–red. We have previously shown that the technique is reliable and rapid. We now show that the UAD is sufficiently uniform metrically that it can be used to derive, from a single set of scaling data, a wide range of color functions, such as the spectral loci of the unique hues, wavelength discrimination, and similarities among very different colored stimuli. The advantage of deriving a UAD is that it requires only a modest amount of participant time to generate the relevant data, which can be re-generated quickly to meet changing viewing conditions.

We have previously shown that color appearance can be specified very reliably by a procedure of direct scaling of hue and saturation (Gordon, Abramov, & Chan, 1994). In that procedure, participants use percentage scales to describe their sensations of the four unique hue sensations of red, yellow, green, and blue, as well as apparent saturation. Participants' responses to the same stimuli remain stable over long periods of time, do not depend on training or past experience with psychophysical procedures and magnitude estimation, and are independent of the participants' linguistic backgrounds.

It has been clear at least since the time of Sir Isaac Newton (see Newton, 1704, p. 90) that color is a response of a sensory system and is not a direct property of the physical stimulus. This means that color appearance can and does change with viewing conditions. It is therefore useful to have a quick and simple procedure for specifying appearance for any given set of conditions, so that any changes in appearance can be examined and compared. Using our scaling procedure, a complete set of color appearance data can usually be obtained in a single session of about 1 h.

Here, we describe how the scaling data can be used to generate a color space representative of the specific viewing conditions. A standard chromaticity diagram, such as the CIE 1931 x, y, z diagram and any of its later transformations, is not adequate for describing appearance. These diagrams are based on single-wavelength additivity mixtures that specify metameric matches. In practice, the spectral distribution of any stimulus is simply scaled by the fixed coefficients of the system to determine the placement of that stimulus on the diagram. Nowhere does appearance enter into the calculation. For example, a stimulus that plots near the spectrum locus close to about 580 nm will usually appear yellowish. However, its appearance can be shifted drastically if it is surrounded by either a reddish or a greenish field, and yet it still plots to exactly the same point on a chromaticity diagram. Moreover, by their very nature, these diagrams specify stimuli independently of their luminances, and yet it has long been known that many aspects of color appearance vary with luminance—such as, for example, the Bezold Brücke phenomenon and wavelength discrimination. Also, the problems with chromaticity diagrams are compounded

I. Abramov, iabramov@brooklyn.cuny.edu

when spectrally broadband stimuli are used, due to failures of luminance additivity (for a review, see Hurvich & Jameson, 1960). Finally, the CIE system is based on matching self-luminous stimuli. It is known that it is difficult to use such matches to specify reflectance samples, and vice versa (Judd, 1951).

Since viewing conditions change, it would be preferable to derive new color spaces that directly specify appearance. Our hue scaling methods can be used to derive such spaces quickly. This, however, is useful only if the color spaces have uniform properties; that is, distances among stimuli indicate the degrees to which the stimuli differ or are similar to each other. We refer to the color space we derive from the hue and saturation scaling data as a *uniform appearance diagram* (UAD), which we have described very briefly in some earlier presentations at conferences (Abramov, Gordon, & Chan, 1990; Chan, Abramov, & Gordon, 1991).

The UAD is two-dimensional, with orthogonal, bipolar, and independent axes of yellow–blue and green–red (see below for details). We now will show in detail that the UAD is metrically uniform and is a valid representation of color space for a specific set of viewing conditions. We will do so by deriving various measures of color vision directly from the UAD and comparing the values with those obtained from the same participants using traditional psychophysical techniques for obtaining those measures.

We first will briefly describe how a UAD is derived from scaling data. We then will examine the following properties of the UAD. We will (1) compare the wavelengths, derived from the UAD, that elicit the unique hue sensations of yellow, green, and blue with those obtained by direct adjustment of a monochromator's wavelengths; (2) examine the size and shape of the error zones associated with participants' responses at different points across the UAD; (3) compare wavelength discrimination estimated from the UAD with similar functions obtained by direct adjustment of wavelength in a bipartite field; (4) examine large-scale uniformity across the UAD by comparing distances among many different stimuli with the degrees to which they appear different; and (5) consider UADs derived from scaling of reflectance samples, as well as the shapes of the appearance curves for stimuli of equal saturation.

Different participants provided the data in the different studies described below; their particular ages, and so on, are given with each data set. All had their color vision screened with the Dvorine pseudo-isochromatic plates and assessed with the Farnsworth Dichotomous Test for Color Blindness (Panel D-15), and the Farnsworth–Munsell 100 Hue Test. All the participants knew the aims of the studies, often serving as experimenters as well as participants.

The studies were approved by the Institutional Review Board of Brooklyn College. All tests were performed at Brooklyn College. All the participants gave informed consent to participate in this study, and the experiments were conducted in accordance with the principles embodied in the Declaration of Helsinki (Code of Ethics of the World Medical Association).

Derivation of a UAD

“The only method we have of knowing anything about the magnitude of the color sensation is by a process of mental estimation.” Moreover, the fact that the observations can be carried out under normal conditions of viewing, using both eyes and without the interposition of any optical devices, adds enormously to the value of the results. (Wright, 1981)

Derivation of a UAD begins with our magnitude estimation technique for assessing color sensations: hue and saturation scaling of stimuli, usually equiluminant, for a particular set of viewing conditions. Full details of the scaling methods can be found in Gordon et al. (1994). The procedure is sufficiently rapid and reliable that enough data can be gathered in a single session to derive the color space (UAD) that represents those sensations.

Since we have described previously the details of how the UAD is derived (Abramov & Gordon, 2005; Abramov et al., 1990; Chan et al., 1991), only a brief outline will be given here. For simplicity, we will deal first with spectral stimuli. On each trial, participants state the percentages of their sensations that were red, yellow, green, or blue, with the proviso that the total be 100%; they then state the degree to which their sensations were saturated, on a scale from 0% to 100% (4+1 category method; Gordon et al., 1994). We prefer percentage scales since participants seem to use them easily. Were we to allow open-ended numeric scaling, we would have to deal with all the problems associated with equating participants' idiosyncratic number ranges.

Before mean hue and saturation functions are computed for each participant, individual data from each trial are arcsine transformed, to allow for boundary restrictions imposed by percentage scales (Gordon et al., 1994). The method used for normalizing variances of proportions is to apply an arcsine transform to each datum before carrying out any other manipulations, such as averaging (Winer, 1971). The specific form that we used was

$$\text{Sensation \%} = \frac{2 \times \arcsine \sqrt{\frac{\text{untransformed sensation \%}}{100}}}{\pi} \times 100.$$

The limits of the scale values transformed in this way are still 0 and 100.

Each individual's hue data, for each wavelength, are also rescaled by the associated saturation value, so that the hues elicited by each stimulus sum to the saturation value. This procedure is appropriate because we have shown that our participants are indeed performing ratio scaling (see Gordon et al., 1994).

Following the concept that hues are represented as two independent, spectrally opponent hue mechanisms (e.g., De Valois & De Valois, 1993; Gordon & Abramov, 2001; Hurvich & Jameson, 1955), the rescaled hue and saturation data can be divided into two curves—green–red and yellow–blue—in which red and green are mutually exclu-

sive, as are yellow and blue. Any minimal overlaps in an individual's mean data between red and green or blue and yellow at the transitions in each opponent mechanism are accounted for by intertrial variability in choice of the pair of terms used to describe a particular hue sensation. For example, no participant ever used red and green together, so that in the vicinity of unique yellow any secondary hue might be either red or green; from trial-to-trial, the precise balance in the opposed spectrally opponent mechanisms can shift slightly, which can tip the sensation, and its description, toward yellow with a little red or yellow with a little green. Interobserver differences account for similar overlaps in group means. For those wavelengths in which there are three hue values for a stimulus, we apportion the lowest value across the remaining two, while still preserving the ratio between the larger values; for example, if the values for a stimulus were 50% yellow, 10% red, and 6% green, the 6% would be divided between yellow and red, so that yellow becomes 55% and red becomes 11% ($50/10 = 5$ and $55/11 = 5$). This resolves the data into precisely two wavelength functions for each participant, from which group means can be computed. These can be plotted on a two-dimensional color space, the UAD, whose axes, green–red and yellow–blue, are orthogonal and bipolar. (Choice of which hue is assigned a negative value is arbitrary.)

Distance of any point from the origin represents its saturation, and its location defines its hue. Since the axes are orthogonal and independent, we have found that the appropriate metric for this space is *city-block*, in which the distance between any two points is measured separately on each axis and the distance is the sum of these two measures. We demonstrated this in an earlier article (Gordon & Abramov, 1988), in which we used multidimensional scaling (Kruskal & Wish, 1978) to reduce a set of hue and saturation scaling data to a space with the least number of dimensions and to compare the *goodness of fit* using Euclidean and city-block metrics. The results were quite clear: The space was defined by orthogonal, bipolar axes of green–red and yellow–blue and had a city-block metric (see also Shepard & Carroll, 1966; Wuerger, Maloney, & Krauskopf, 1995).

The spectral range of the stimuli used in any given experiment is an issue. For most of what we present here, the stimuli were monochromatic stimuli spaced in regular steps of 10 nm across the spectrum. We have tested, however, the effects of restricting or varying the spectral range: In an earlier study (Gordon et al., 1994), we used the same number of stimuli, but spaced in 5-nm steps and restricted to the beginning and end of the spectrum, omitting the middle wavelengths. The mean functions from the restricted stimulus set were indistinguishable from their counterparts when the stimuli were spaced uniformly across the spectrum. Finally, later in this article, we will show a similar effect when stimuli are unequally spaced (see Figure 5).

In the rest of this article, we will show that a UAD for a single set of viewing conditions not only contains sufficient information to describe color appearance, but also allows derivation of a range of other color functions that are usually obtained using more classical psychophysical

techniques. These derivations require that the UAD be metrically uniform. In some cases, the requirement is more stringent than others. We will begin with the least stringent and will progress to more stringent cases, in each of which we will compare the functions derived from a UAD with functions generated by the same participants using conventional methods.

UADs: Wavelengths of Unique Hues

One of the less demanding derivations of other color functions from a UAD is to locate the spectral loci of the unique hues. These loci (e.g., the precise wavelength that appears purely yellow) are theoretically important, since they represent the null points of particular opponent hue mechanisms (e.g., in the case of unique yellow, the green–red mechanism; Abramov & Gordon, 2005; Gordon & Abramov, 2001; Hurvich & Jameson, 1955).

Figure 1A shows mean data from a group of 9 participants (7 of them male, 2 female; mean age, 22.8 years; range, 19–30 years). Scaling methods and data manipulations were as described above. (Note that all individual responses to any stimulus were arcsine transformed, as described above, before any form of averaging or other data manipulation occurred.) Stimuli were seen foveally in a Maxwellian view optical system with grating monochromators controlling stimulus wavelengths: 440–660 nm, in 10-nm steps, against a dark background, 500-msec duration, 20-sec interstimulus interval, 1° subtense, 100-Td retinal illuminance; stimulus area outlined by four dim achromatic points that were continuously visible. Stimulus wavelengths were presented in blocks, in random order within a block; all the stimuli were presented once in each block; the first block was considered practice and was not included in analyses; data from the next six blocks were analyzed, so that an individual's results were means of six repetitions.

The UAD corresponding to the group data in Figure 1A is shown in Figure 1B. The data points are the scaled values at each 10-nm point. To reduce noise in the UAD, we apply a running average in which each data point is averaged with half the value of its immediate neighbors, and then a smoothed cubic spline is fitted to these points and evaluated at 1-nm intervals (Chan et al., 1991). Fitting such a continuous function to the data allows us to derive the wavelengths of a variety of other functions whose wavelengths might not have been included in our test set. Fitting this function does not introduce any major distortion. We always check this, for each individual participant, by plotting the smoothed spline function together with the original individual's arcsine-transformed data, and the agreement is always excellent. Figure 1B shows the same agreement for the group data in Figure 1A. Much of the rest of this article will show that functions derived from our UADs agree very well with those derived from more classical techniques.

The wavelengths of unique hues can be ascertained from Figure 1B, by interpolating on the smoothed function to find the wavelength at which it crosses one of the axes. For example, unique blue must lie between 460 and 470 nm for the given viewing conditions; the exact value is found by interpolation.

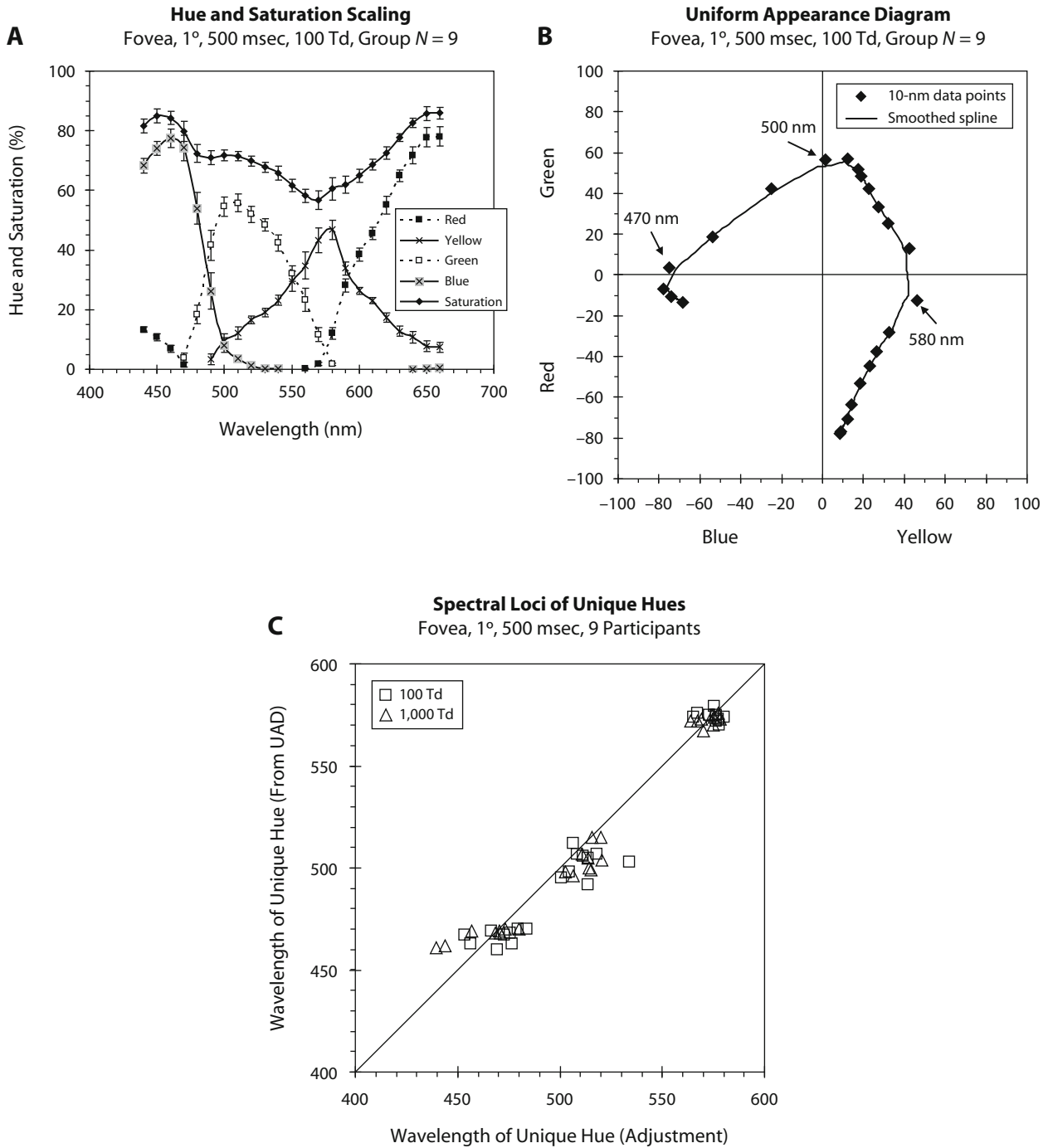


Figure 1. Magnitude estimation of color appearance from participants who viewed (Maxwellian view) monochromatic lights against a dark background. (A) Hue and saturation functions: group means ± 1 SEM. (B) Data in panel A replotted in a 2-D color space, or uniform appearance diagram (UAD); the curve is a smoothed cubic spline fitted to the data. (C) Comparison of the spectral loci of the unique hue sensations of yellow, green, and blue, derived by interpolation on the spline function in panel B and from direct adjustment of the wavelength of a monochromator to locate each of the unique hues. Data points are for the individual participants and are for two retinal illuminances, 100 and 1,000 Td.

The wavelengths for the unique hues of blue, green, and yellow were obtained for each participant and are shown in Figure 1C, using the y-axis. The same participants repeated the entire experiment at 1,000 Td, and these data are also included. To validate these interpolations, the

same participants located the wavelengths of the unique hues by a direct adjustment procedure at each of the two retinal illuminances.

The same apparatus was used as for the hue and saturation scaling. The procedure was as follows. A 0.5-sec

warning buzzer was followed by a 0.5-sec dark period preceding the 500-msec flash; this was followed by a 3.5-sec dark period before the cycle was repeated. The participant controlled a motor that could change the wavelength setting of the monochromator. Adjustments were made during the dark period, and the participants continued until satisfied that they had found the unique hue. To avoid biases due to choice of starting points, there were eight starting wavelengths, in 5-nm steps, associated with each unique hue: 455–495 for blue, 485–525 for green, and 555–595 for yellow. The order of the hues was randomized across participants, with two hues run at a given illuminance per session. Within a session, a hue was run in three blocks with randomized starting wavelengths, for a total of 24 determinations for each unique hue.

The individual participants' adjustment data are shown in Figure 1C, using the x -axis. Clearly, the agreement between the two methods is excellent. The difference is that obtaining the data in the UAD requires far less time: Two sessions, one for 100 Td and one for 1,000 Td, each lasting about 1 h, were sufficient. Four sessions were needed to get comparable data by adjustment.

UAD: Variance Zones

One indicator of the metric uniformity of a UAD is that the error zones associated with each position be the same size and shape—that is, that there be no bias in favor of any one direction through this color space. This is the logic that led the C.I.E. to search for transformations of the original 1931 x, y, z chromaticity diagram into diagrams such as the 1976 UCS u', v' chromaticity diagram (Wyszecki & Stiles, 1982). In the former, the variance ellipses were far from circular and varied grossly in size across the diagram, whereas in the latter, they were more uniform but still far from truly uniform (Ohno, 2006).

Ideally, the size and shape of our variance zones should be examined across the entirety of a color space. However, in the great majority of our studies, we have used monochromatic stimuli, so that the variance zones we can derive are usually associated only with the spectrum locus. For example, in Figure 2A, we show a UAD for a group of 16 young participants (7 of them male, 9 female; age, 20–27 years), who viewed rear-projected (Newtonian view, using natural pupil) monochromatic light flashes (500 msec, 27 Td, 1°, fovea, dark surround, 20-sec inter-trial interval); scaling procedures were our standard 4+1 categories method (see above). Each participant's mean hue and saturation values, and associated variances, were computed for the four hue categories. The individual participant's variances for green–red and for yellow–blue were averaged to give the group mean variances along each of the axes of the UAD. In the figure, we plot the square roots of the variances about the data points along the two axes. (For clarity, we show only the error zones for alternate data points.) There do not seem to be large differences in the extents of the error zones, nor are there obvious asymmetries associated with the axes: The error zones are approximately the same size at all points along the spectrum locus.

In another study, we explored the effects on color appearance of desaturating the spectral stimuli by adding fixed amounts of equal-energy white light (a good approximation to a truly achromatic stimulus) while keeping the total retinal illuminance the same. The resulting color spaces from 25% to 100% colorimetric purity are shown in Figure 2B. As was expected, reducing stimulus purity restricts the color space, drawing all points closer to the diagram's origin, which is the completely achromatic point. These data allow us to explore the error zones within the UAD away from the usual spectrum locus. In Figure 2C, we show the same sort of error zones as those in Figure 2A, but for the 100% and 25% stimulus sets. Again, the error zones all seem roughly equal in extent, and although their sizes do show greater variability (there were far fewer participants in this study), there do not seem to be any systematic biases according to locus.

To examine more closely the “equality” of the error zones across the UAD, we have computed the aspect ratios of the error bars for each of the data points in Figure 2B—that is, the length of the error bar along the yellow–blue axis to the corresponding length of the error bar along the green–red axis. The results are shown in Figure 2D. Although the ratios are clustered about a value of 1.0 (implying equality), there is clearly a trend across wavelength: Along the yellow–blue axis, the error zones are wider at longer wavelengths and narrower at shorter wavelengths. To test whether this was simply noise due to the relatively small sample of participants, we performed the same analyses on the data shown in Figure 2A, for which the sample was much larger. The results are plotted on exactly the same coordinates in Figure 2E. Although the deviations from unity are smaller, the same trend across wavelength is evident.

Interestingly, the same trend can be found in the standard CIE chromaticity diagrams, including the 1960 u, v and 1976, u', v' spaces: In these spaces, the error ellipses are more nearly circular, but there is still a considerable variation in their aspect ratios, so that near the longer wavelengths, the horizontal axis is longer, and near the shorter wavelengths, the vertical axis is longer (see Ohno, 2006; Wyszecki & Stiles, 1982). This variation would produce the same trend as that in our UAD. We therefore conclude that the trend in Figures 2D and 2E is not an artifact of our scaling procedures. The reasons for this trend in the shapes of the error zones are unclear.

In Figure 2F, we plot against each other the mean variances associated with each axis for each data point in Figure 2B; there does not appear to be any correlation between the variances in the two axes of the UAD, which is the expectation if these are independent dimensions.

Our scaling studies, in which stimuli that varied in colorimetric purity have been used have provided much of the data attesting to the uniformity of our UAD (Figures 2C, 2D, 2E, and 2F). We present in Figure 2G additional analyses of those data to verify their validity. We selected a series of hues including the unique hues, the intermediate hues (e.g., 50% green/50% yellow), and hues midway between the uniques and the intermediates

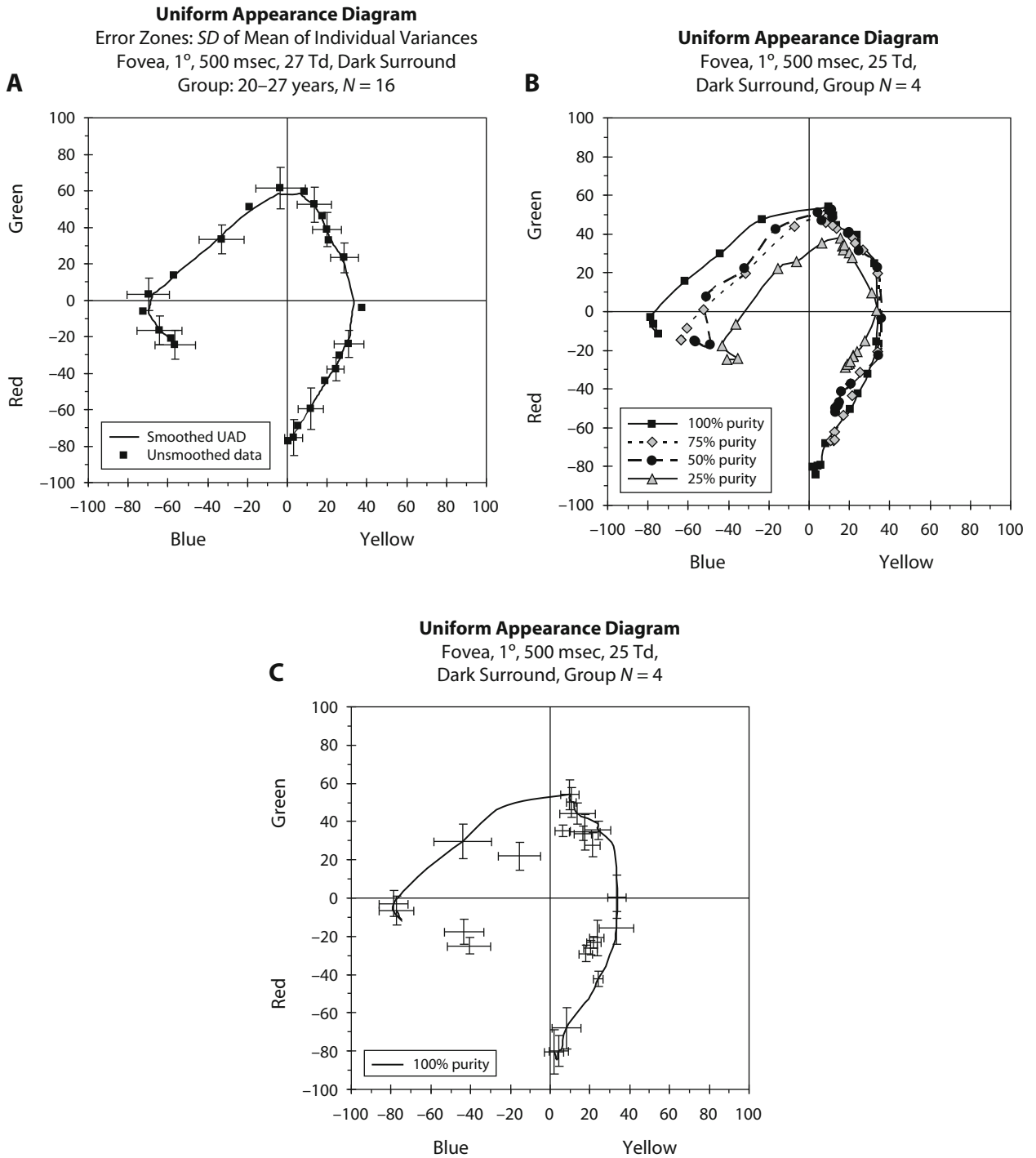


Figure 2. Variance zones. (A) Uniform appearance diagram (UAD) showing results of hue and saturation scaling by a group of participants viewing monochromatic lights. The square roots of the variances along each color axis are plotted for alternate stimulus points. (B) UAD showing results of hue and saturation scaling by a different group of participants viewing monochromatic lights under similar conditions, except that equal-energy white light was added to vary the purity of the stimuli systematically from 25% to 100%. (C) As in panel A, square roots of the variances along each color axis are plotted for alternate stimulus points for the 25% and 100% purity conditions.

(e.g., 75% green/25% yellow). For each we derived, from the smoothed spline function, the corresponding wavelengths. Clearly, the wavelengths eliciting a particular hue are not constant. The shifts are as would be expected from

the Abney effect (see, e.g., Newhall, Nickerson, & Judd, 1943). Consider, for example, unique blue: At 100% purity, it is elicited by a wavelength of approximately 460 nm; as purity decreases, the wavelength increases. Perceptually,

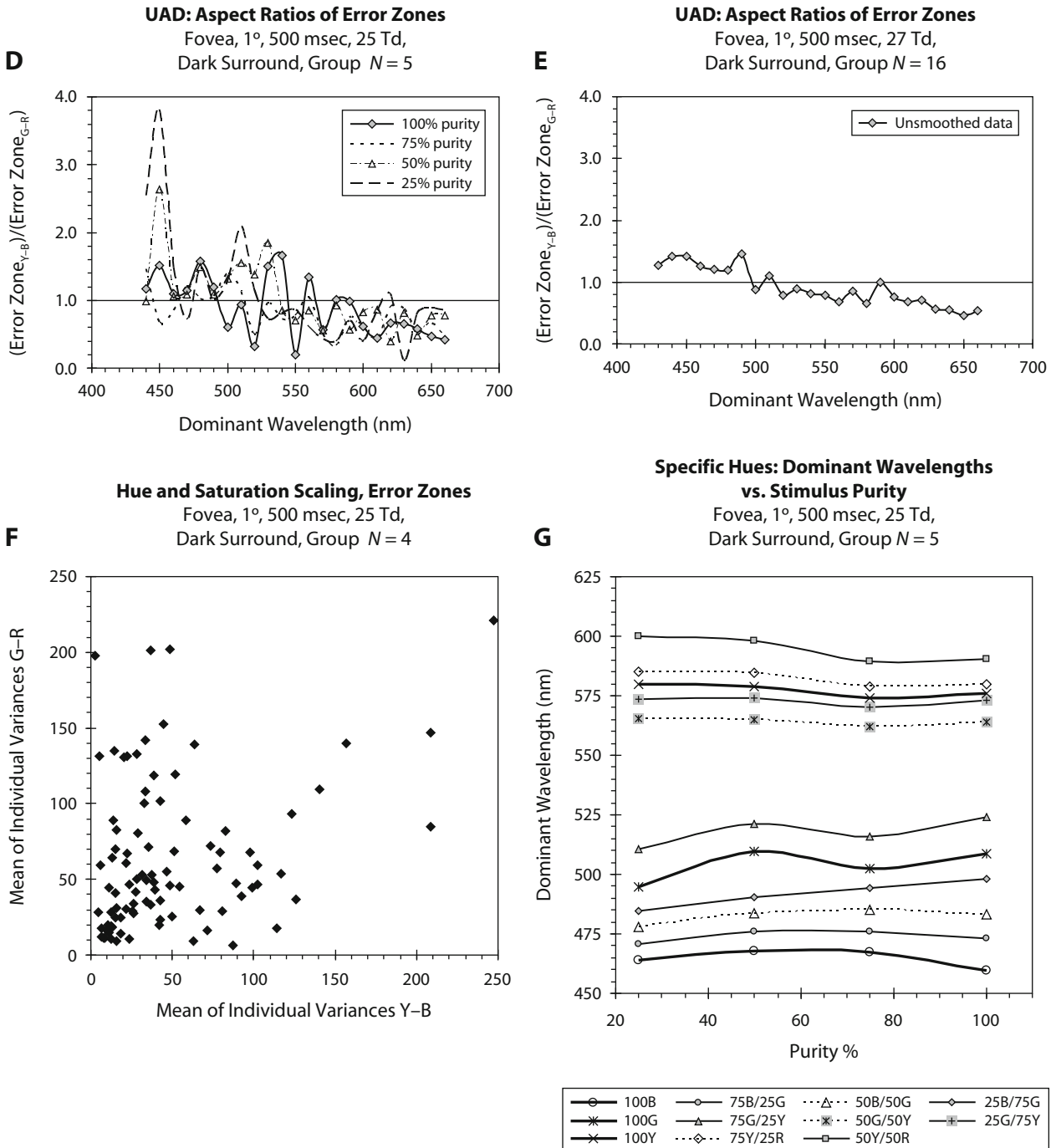


Figure 2 (continued). Variance zones. (D) Aspect ratios of error zones for all the data points in panel B. For each datum, the length of the error zone along the yellow–blue (Y–B) axis was divided by the associated length along the green–red (G–R) axis. (E) Aspect ratios of error zones for all the data points shown in panel A. Same conventions as in panel D. (F) Mean variances in the Y–B axis associated with each of the data points shown in panel B versus mean variances for the same data points in the G–R axis. (G) Abney effect. Variations in dominant wavelengths eliciting specific hue sensations as colorimetric purity is changed (derived from the UAD in panel B).

as the purity of a short wavelength is decreased, its appearance becomes tinged with red; to maintain a sensation of unique blue, a longer wavelength must be used that elicits a small amount of greenness that negates the small amount of redness.

Small-Scale Uniformity: Wavelength Discrimination

Wavelength discrimination measures the minimally detectable change in color appearance due to a change in stimulus wavelength at each point along the spectrum.

Usually this is done with a bipartite field, in which wavelength of one half-field is increased or decreased to find the positive and negative increments needed to elicit just-noticeable differences, and the average of these is taken as the discrimination value for the given spectral locus. Comparable discrimination functions can be obtained from measures of color appearance (Graham, Turner, Holland, Bradley, & Burdeshaw, 1976; Hurvich & Jameson, 1955; Smith, 1971; however, most of these studies have used color naming, which does not provide the continuous measure of color appearance given by our scaling method). We can derive a comparable function from the UAD by interpolating to find the wavelengths that would be required to elicit some small criterion change in color appearance. Comparing the two functions would test uniformity of the UAD for small distances about each spectral point.

Three color-normal male participants (age, 18–34 years) provided the data. All the stimuli were seen foveally in Maxwellian view (1° , 100 Td, 500 msec, dark surround). Traditional discrimination functions were obtained using a bipartite field, one side of which was set to a fixed wavelength; a participant adjusted the wavelength of the other side, via a motor controlling the wavelength of one of the optical system's grating monochromators. We used 21 wavelengths from 450 to 650 nm in 10-nm steps, randomized within a block, with four blocks of 42 trials per session, for a total of four sessions. Each trial began with both sides of the field set to the same wavelength; a trial was designated, randomly, as positive or negative, indicating the direction in which the participant was to adjust the wavelength. The total repeats for each spectral point was eight. Adjustments were made during the dark periods after flashes, and a participant could continue adjusting to and fro until satisfied.

In Figure 3A, we show the group average for the adjustment data, together with the group *SEMs*. For comparison, we include a standard discrimination curve, obtained in similar fashion, from the literature (Bedford & Wyszecki, 1958).

For comparison with these curves, we derived curves from each participant's UAD for the same spectral stimuli. The same wavelengths were scaled, using our standard procedures, except that for added precision, each stimulus was seen a total of eight times. Starting with the UAD values for each spectral stimulus, we calculated the wavelength that would have been associated with a small criterion shift in appearance in both positive and negative directions and averaged these to obtain the discrimination value for that spectral locus. All measurements were made using the city-block metric. Since there is no necessary value that relates magnitude of change in color appearance to just-noticeable difference in wavelength, we chose the value of the criterion change in color appearance that gave the best fit (least-squares rule) to each individual participant's adjustment data. The respective UAD values for best fit for the participants were 2.0, 2.8, and 4.2. The group means for these UAD-derived curves, and their *SEMs*, are also shown in Figure 3A. The difference between the procedures is that using the UAD requires only one session, rather than four lengthy sessions.

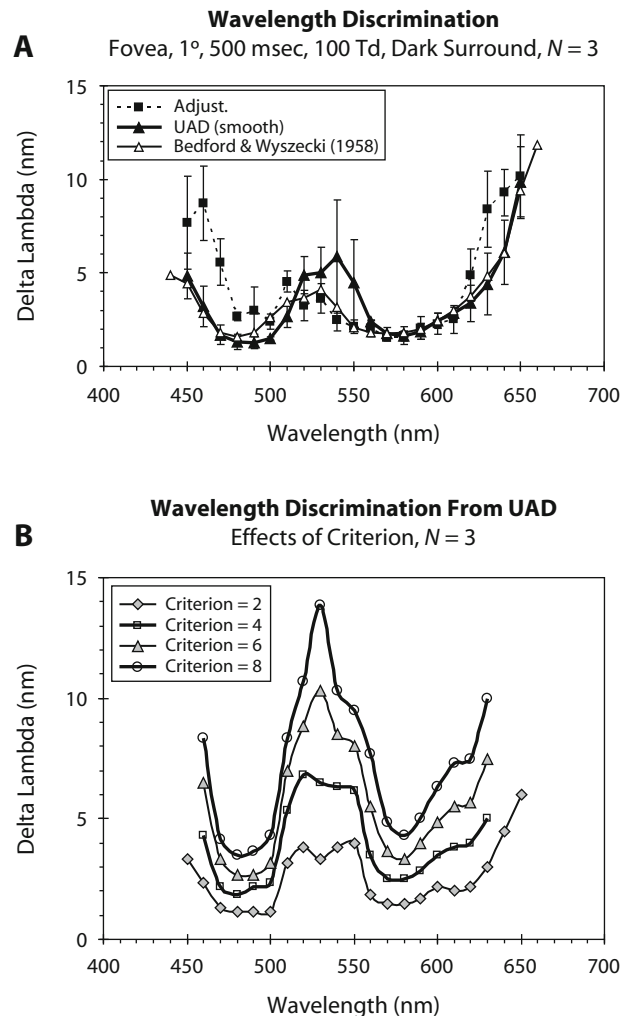


Figure 3. Wavelength discrimination. (A) Discrimination functions for 3 participants. One curve is the mean, ± 1 SEM, derived from each participant's UAD from hue and saturation scaling of the monochromatic light flashes. The second curve from the same participants is the mean, ± 1 SEM, of their individual discrimination curves derived from adjusting the wavelength of a bipartite field to produce just-noticeable differences (JNDs). For comparison with our data, we include a canonical curve from the literature for adjustment of a bipartite field at the same retinal illuminance as in our case (Bedford & Wyszecki, 1958). (B) Mean of individual discrimination functions (delta lambda in nm) derived from their UADs, using different criteria for a JND.

As we have noted, the problem lies in choosing a criterion shift in the UAD values to represent a just-noticeable difference and then evaluating its corresponding shift in wavelength. We argue that the essence of any derived wavelength discrimination function is its ability to capture the features of the classically measured discrimination function. Does choice of criterion change in appearance affect the shape of the derived discrimination function? Using the UADs for each of the 3 observers in Figure 3A, we derived discrimination functions for criterion changes over a range of plausible values: 2, 4, 6, and 8 (on a typical UAD—e.g., Figure 1B—the city-block UAD distances between stimuli 10 nm apart range from approximately 2 to 25).

We show the group means of these functions, for each criterion value in Figure 3B. Ideally, they should differ only by a scale factor: Thus, the values using a criterion of 8 should be twice those for the criterion of 4, and so on. We computed the ratios of wavelength shifts associated with each criterion value and compared them with the ideal values. The ideal scale factors by which these curves should differ, when compared with the curve for a criterion of 8, are 4, 2, 1.33, and 1. The means from our measured values are 3.21, 1.83, 1.29, and, of course, 1. (The worst case is for the smallest change in appearance, a criterion shift of 2; we interpret this as a *floor* effect.) The ratios we obtained closely approximate our expectation that, for a limited range of criterion values, the derived discrimination functions should differ only by scale factors. We have since settled on a criterion shift of 4 as a conservative estimate of the change in appearance required to produce a just-noticeable difference and use this in all our other work.

For this article, we did not use our usual value of 4 but found the best criterion for each individual in order to illustrate more clearly how good a fit can be obtained. Since the adjustment procedure is very tedious and time consuming, we tested only a few observers; as was expected, there was considerable variation among them. (All published adjustment data show such variation when individual data are shown; see, e.g., Bedford & Wyszecki, 1958; Walraven, 1962.)

Large-Scale Uniformity: Color Differences

In order to describe a color space for a given set of viewing conditions completely, a UAD should also be able to specify precisely how different stimuli appear when they are far apart in this space—that is, when they are perceptually quite different. This is a problem that is of some importance in many industrial settings (Pointer & Attridge, 1997). To achieve this, the UAD must be metrically uniform over large distances, as is suggested by the data shown in Figure 2C.

Five participants (3 of them male, 2 female; age, 16–25 years) used our standard procedures to scale the appearances of spectral stimuli from 440 to 660 nm (foveal, Maxwellian view, 35 Td, 500 msec, dark surround). The resulting UAD for the group is shown in Figure 4A. For this study, each stimulus was repeated a total of eight times across two sessions.

Ten of the stimuli, spaced across the spectrum, were chosen for the next phase. On each trial, pairs of these stimuli were presented: Following a warning buzzer, the first was shown for 500 msec; then, 3 sec later, the second was shown for 500 msec; after a 15-sec intertrial interval, the next pair was presented; and so on. All pairwise combinations of the 10 stimuli were used, including reversing the orders of any given pair, for a total of 100 trials per session. In addition, a session began with a random selection of 10 pairs for practice, which were not included in the analyses. In the dark period following presentation of the pair of stimuli, the participants assigned a percentage value to the hue difference they perceived between the two stimuli; a small difference was denoted by a small value,

and a large difference was denoted by a large value; they were to attend only to the perceived difference in color appearance, and not to any differences in brightness.

Since the difference-scaling responses also used percentage scales, like hue and saturation scaling, they too were arcsine transformed before deriving means for individual participants. The city-block distances between each pair of stimuli were derived from the smoothed UADs, and were compared with the perceived differences obtained from direct scaling. Figure 4B shows these comparisons for the 5 individual participants, together with the group mean.

All individuals show a clear linear relationship between degree of similarity, or difference, between stimuli and their distances across the UAD, regardless of the magnitudes of these distances. Each individual's data are well fit by a straight line. The *R* values of the correlations range from .88 to .97, with a mean of .94; the slopes vary from 1.26 to 2.07, with a mean of 1.84. Thus, individuals differ from each other only by a relatively small scale factor. A related question is whether there are differences when the data cross from one UAD quadrant to another (i.e., change from one pair of hue categories to another). Figure 4C shows data separated into series that cross from one quadrant to another, including diagonally. There are no systematic differences. We argue that this constitutes strong evidence that the UAD is uniform across its entire extent.

Reflectance Samples of Equal Saturation and Lightness

If our method is to be totally general, it should apply just as well to reflectance samples (real objects) as to lights. We tested this by having participants scale the color appearance of a standard set of reflectance samples (Munsell papers) taken from the Farnsworth–Munsell 100-Hue Test. In this test, all the caps are equal in chroma (saturation) and value (lightness) when viewed under the appropriate illuminant; only the hue varies among the caps in small, equal shifts.

We chose 31 of the caps from the original test; they were not equally spaced but were chosen for purposes of another, unrelated experiment for which we needed descriptions of their color appearance. Figure 5A shows the specific caps chosen (stimuli viewed foveally; subtense, 1°; duration, 500 msec; 80 Td; intertrial interval, 20 sec). The stimuli were presented, one at a time, on an easel inside a large enclosure (gray, 20% reflectance) and were illuminated by tungsten sources filtered by mired shift filters (Roscoe Cinegel) to provide an illuminant of approximately 6,000 K. Using our scanning spectrophotometer/radiometer, we measured the spectrum of each stimulus as viewed by a participant. From these data, we computed dominant wavelengths with respect to equal-energy white (a close approximation to the achromatic point). Each stimulus was placed behind a large, circular-aperture, mechanical shutter. The participant's head was stabilized with a chin and forehead rest. All scaling procedures (4+1) were the same as those described for all the previous studies. The stimuli were presented in blocks, in random order within a block; the first block was considered practice and

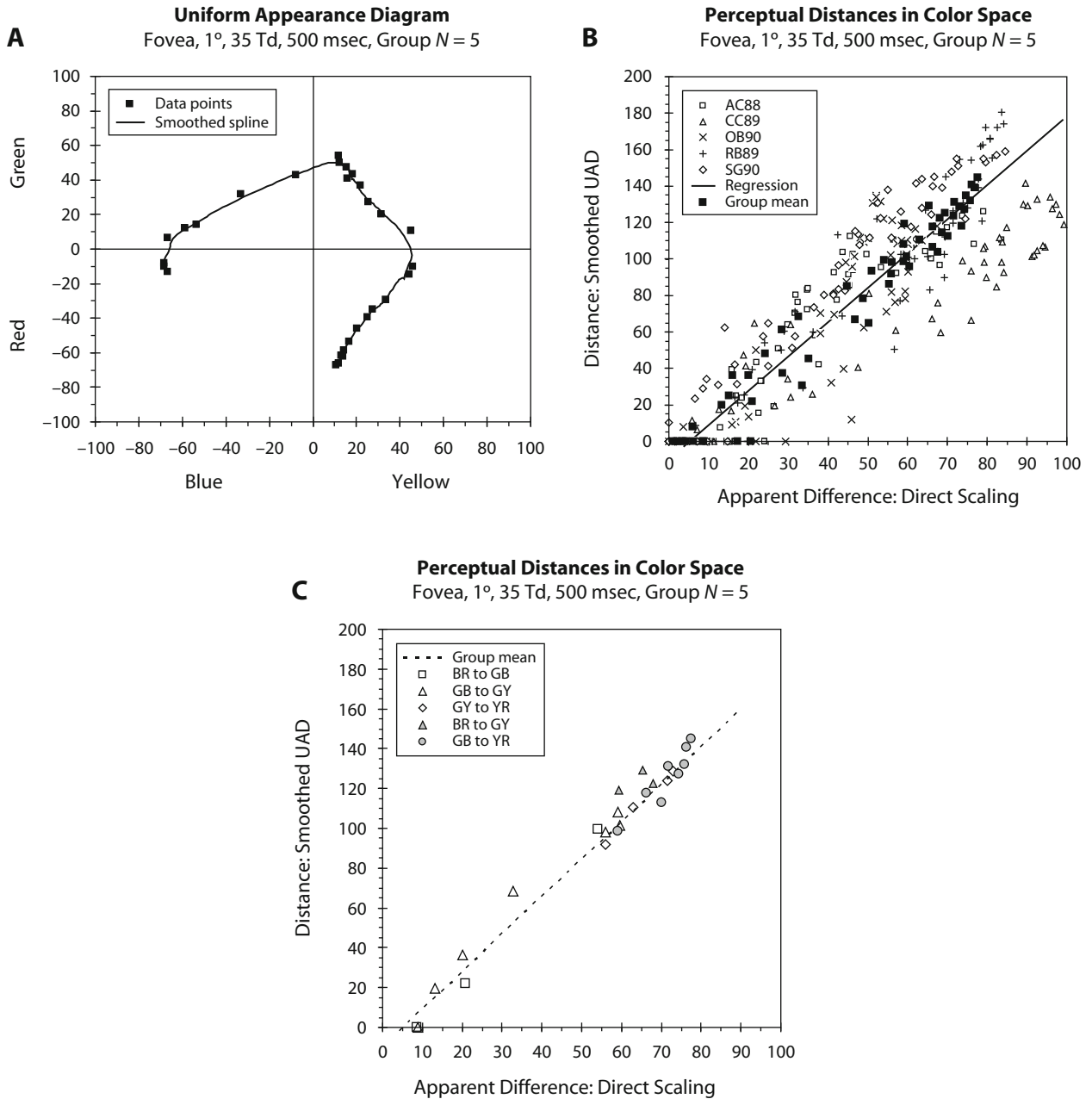


Figure 4. (A) Uniform appearance diagram (UAD) showing results of hue and saturation scaling by a group of participants viewing monochromatic lights. (B) Pairwise comparisons of 10 stimuli spaced across the spectrum. Participants scaled the apparent difference between the stimuli in each pair, and these values are plotted along the x-axis. For comparison, the distances between the same stimuli are derived from the UAD and plotted on the y-axis. Data are shown for each individual participant, together with the group mean. (C) Distances between stimuli located in different quadrants of the UAD, derived from the group mean shown in Figure 4B.

was not included in analyses; data from the next six blocks were analyzed.

In Figure 5B, we show the group mean hue and saturation functions, together with their SEM bars, using dominant wavelengths of the stimuli for the x-axis. These data, from relatively broadband reflectance samples, are remarkably similar to the functions we have shown for monochromatic lights. We also point out that, in some parts of the spectrum, the dominant wavelengths of the

Munsell caps are more closely spaced than are the wavelengths in the studies above. In particular, we note that the saturation function approximates a horizontal line, which is as it should be, given that all the caps are of equal chroma. Also, despite the close spacing of some of the stimuli, there are no reversals in the hue scaling functions. Furthermore, in the 100-Hue Test, the caps are viewed simultaneously and continuously, which facilitates comparisons of their hues; in our case, they were viewed

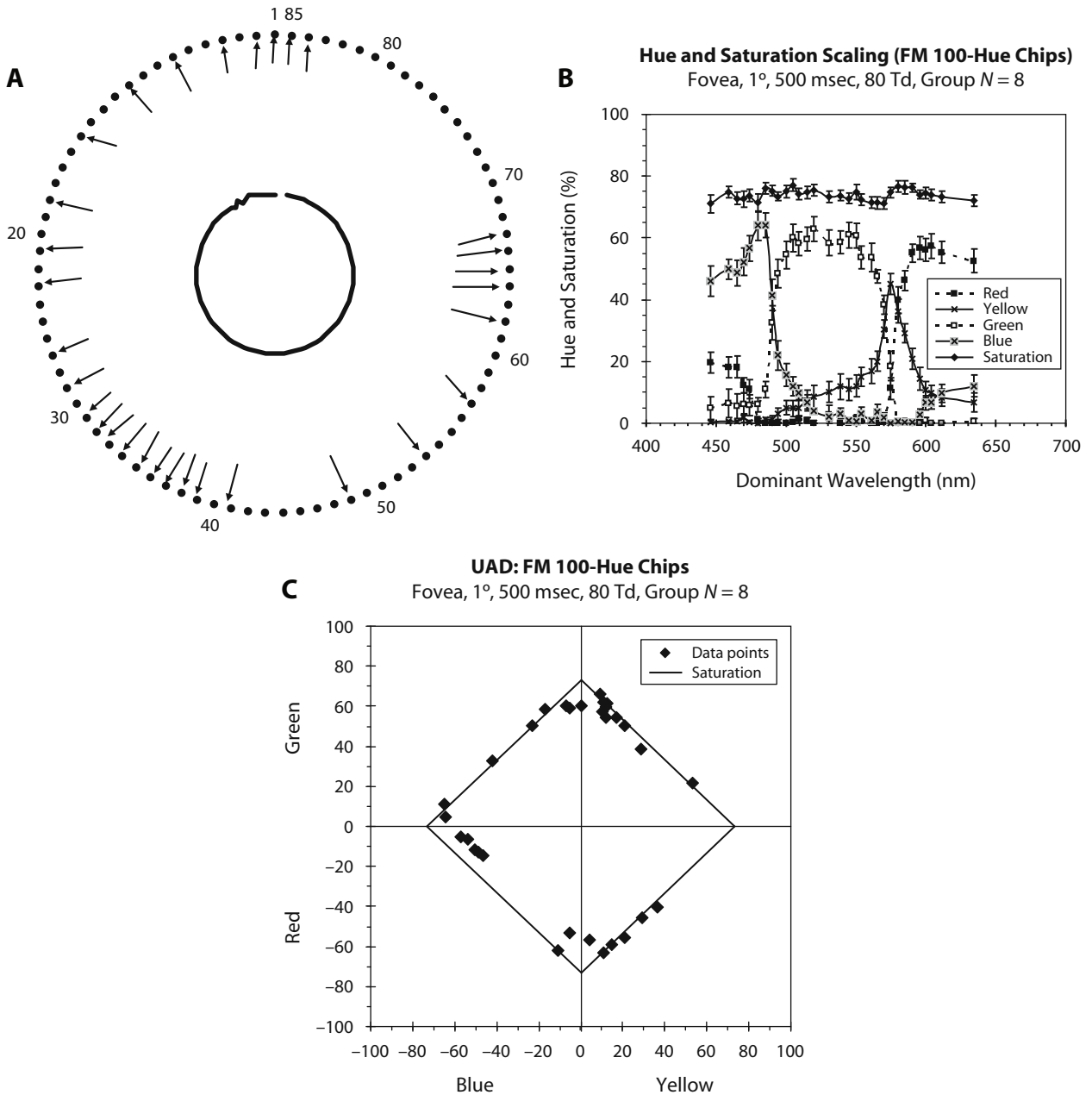


Figure 5. Scaling color appearance of Munsell chips. (A) Caps chosen from Farnsworth–Munsell (FM) 100-Hue Test for scaling. (B) Group mean hue and saturation functions ± 1 SEM. (C) Hue and saturation data replotted on UAD; lines are the overall mean saturation function from Figure 5B.

one at a time, so that the participants had to make absolute judgments of appearance.

Figure 5C shows the UAD of these data. In this case, since there is no real variation in saturation, all the data points plot along the diamond-shaped locus predicted from our city-block metric. The lines in this figure are exactly the overall mean saturation computed from Figure 5B.

Conclusions

Hue and saturation scaling of the color appearance of stimuli can be presented usefully on a two-dimensional

diagram, the UAD, which we have shown in several ways above to be a uniform space, with a city-block metric. The first advantage of our approach is that, as viewing conditions change, it is easier to comprehend the sensory changes from changes in the color space than it is from the more complicated sets of direct scaling data. For example, the effects of desaturating spectral stimuli are easily demonstrated by superimposing the UADs (see Figure 2B) but would require several sets of hue and saturation curves.

Perhaps of greater importance is the amount of additional data that can be derived from any one UAD. A

single set of scaling data, requiring, at most, one or two experimental sessions, can yield functions that, together, would require many more experimental sessions. For example, the wavelength discrimination data (Figure 3) obtained by direct adjustment of stimulus wavelength to obtain just-noticeable differences called for twice as many sessions as did the time it took to obtain the scaling data for the UAD. And the adjustment sessions yielded only the discrimination data, whereas the same UAD could be used to derive a variety of other functions.

This is not to denigrate the traditional methods. Rather, our point is that our approach provides a complementary look at these functions. The usual chromaticity diagrams are extremely useful for specifying stimuli: All that is needed is the appropriately weighted spectral distribution of a stimulus to locate that stimulus on the diagram; and if two stimuli, even of different physical spectra, happen to plot to the same point, then for the visual system, they are equivalent (they are metamers) and produce the same sensation, whatever that might be. The problem is that color appearance can and does change with viewing conditions and the state of the viewer. Hue and saturation scaling describes sensory appearance under each specific set of conditions.

Moreover, the time savings allow one to test a larger number of participants than is the usual case with more time-consuming procedures. This may be a particularly useful feature in the light of the growing awareness that human color vision can be quite variable, in terms of cone polymorphism, sex differences, and variations in cone ratios (e.g., Neitz & Neitz, 1998; Roorda & Williams, 1999). Furthermore, hue scaling may be the optimal procedure for describing color constancy in 3-D scenes under different illuminants (Schultz, Doerschner, & Maloney, 2006).

Finally, we point out that our scaling method can be applied to any chromatic stimulus. We have shown above that it can be used with monochromatic stimuli as well as with broadband reflectance stimuli. There is no reason why it cannot be extended to any complex stimulus. Any light stimulus can be matched by an additive mix of three primaries; the resulting metamer exerts the identical impact on the visual system and locates the stimulus on a specific point on a standard chromaticity diagram. As we have shown above, our scaling methods work well with stimuli located well away from the spectrum locus and covering the gamut of a chromaticity diagram.

AUTHOR NOTE

This research was supported in part by grants to I.A. and J.G. from the PSC-BHE Research Awards Program, City University of New York. Some of the equipment was bought with funds from a New York State Higher Education Advanced Technology Program Grant to Brooklyn College. Correspondence concerning this article should be addressed to I. Abramov, Department of Psychology, Brooklyn College of the City University of New York, 2900 Bedford Avenue, Brooklyn, NY 11210 (e-mail: iabramov@brooklyn.cuny.edu).

REFERENCES

ABRAMOV, I., & GORDON, J. (2005) Seeing unique hues. *Journal of the Optical Society of America A*, **22**, 2143-2153.
 ABRAMOV, I., GORDON, J., & CHAN, H. (1990). Using hue scaling to

- specify color appearance. *Proceedings of the Society of Photo-Optical Instrumentation Engineers*, **1250**, 40-51.
 BEDFORD, R. E., & WYSZECKI, G. W. (1958). Wavelength discrimination for point sources. *Journal of the Optical Society of America*, **48**, 406-411.
 CHAN, H., ABRAMOV, I., & GORDON, J. (1991). Large and small color differences: Predicting them from hue scaling. *Proceedings of the Society of Photo-Optical Instrumentation Engineers*, **1453**, 381-389.
 DE VALOIS, R. L., & DE VALOIS, K. K. (1993). A multistage color model. *Vision Research*, **33**, 1053-1065.
 GORDON, J., & ABRAMOV, I. (1988). Scaling procedures for specifying color appearance. *Color Research & Application*, **13**, 146-152.
 GORDON, J., & ABRAMOV, I. (2001). Color vision. In E. B. Goldstein (Ed.), *The Blackwell handbook of perception* (pp. 92-127). Oxford: Blackwell.
 GORDON, J., ABRAMOV, I., & CHAN, H. (1994). Describing color appearance: Hue and saturation scaling. *Perception & Psychophysics*, **56**, 27-41.
 GRAHAM, B. V., TURNER, M. E., JR., HOLLAND, R., BRADLEY, E. L., & BURDESHAW, J. A. (1976). Wavelength discrimination derived from color naming. *Vision Research*, **16**, 559-562.
 HURVICH, L. M., & JAMESON, D. (1955). Some quantitative aspects of an opponent-colors theory: II. Brightness, saturation, and hue in normal and dichromatic vision. *Journal of the Optical Society of America*, **45**, 602-616.
 HURVICH, L. M., & JAMESON, D. (1960). Color vision. *Annual Review of Psychology*, **11**, 99-130.
 JUDD, D. B. (1951). Report of the U.S. secretariat committee on colorimetry and artificial daylight. In *Proceedings of the 12th Session of the CIE*. Stockholm: Bureau Central de la CIE.
 KRUSKAL, J. B., & WISH, M. (1978). *Multidimensional scaling* (Sage University Paper Series on Quantitative Applications in the Social Sciences, 07-001). Beverly Hills, CA: Sage.
 NEITZ, M., & NEITZ, J. (1998). Molecular genetics and the biological basis of color vision. In W. G. K. Backhaus, R. Kliegl, & J. S. Werner (Eds.), *Color vision: Perspectives from different disciplines* (pp. 101-119). Berlin: Walter de Gruyter.
 NEWHALL, S. M., NICKERSON, D., & JUDD, D. B. (1943). Final report of the O.S.A. Subcommittee on the Spacing of Munsell Colors. *Journal of the Optical Society of America*, **33**, 385-418.
 NEWTON, I. (1704). *Opticks: Or a treatise of the reflexions, refractions, inflexions and colours of light*. London: Sam. Smith and Benj. Walford.
 OHNO, Y. (2006). Basic concepts in photometry, radiometry and colorimetry. In J. P. Dakin & R. G. W. Brown (Eds.), *Handbook of optoelectronics* (pp. 287-305). Boca Raton, FL: CRC Press.
 POINTER, M. R., & ATTRIDGE, G. G. (1997). Some aspects of the visual scaling of large color differences. *Color Research & Application*, **22**, 298-307.
 ROORDA, A., & WILLIAMS, D. A. (1999). The arrangement of the three cone classes in the living human eye. *Nature*, **397**, 520-522.
 SCHULTZ, S., DOERSCHNER, K., & MALONEY, L. T. (2006). Color constancy and hue scaling. *Journal of Vision*, **6**, 1102-1116.
 SHEPARD, R. N., & CARROLL, J. D. (1966). Parametric representation of non-linear data structures. In P. R. Krishnaiah (Ed.), *Multivariate analysis: Proceedings of an international symposium* (pp. 569-592). New York: Academic Press.
 SMITH, D. P. (1971). Derivation of wavelength discrimination from colour-naming data. *Vision Research*, **11**, 739-742.
 WALRAVEN, P. L. (1962). *On the mechanisms of colour vision*. Soesterberg, The Netherlands: Institute for Perception RVO-TNO.
 WINER, B. J. (1971). *Statistical principles in experimental design*. New York: McGraw-Hill.
 WRIGHT, W. D. (1981). Why and how chromatic adaptation has been studied. *Color Research & Application*, **6**, 147-152.
 WUERGER, S. M., MALONEY, L. T., & KRAUSKOPF, J. (1995). Proximity judgments in color space: Tests of a Euclidean color geometry. *Vision Research*, **35**, 827-835.
 WYSZECKI, G., & STILES, W. S. (1982). *Color science: Concepts and methods, quantitative data and formulae*. New York: Wiley.

(Manuscript received April 23, 2008;
 revision accepted for publication October 28, 2008.)